DOCUMENT RESUME

ED 287 842 SP 029 496

AUTHOR Wingate, James G.; Bowers, Fred
TITLE Appraising Teacher Performance: A Quantitative Approach.
PUB DATE Sep 87
NOTE 21p.; Some tables contain small print.
PUB TYPE Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; *Evaluation Methods; *Models; Outcomes of Education; Secondary Education; Student Reaction; *Teacher Behavior; *Teacher Effectiveness; *Teacher Evaluation; Teacher Student Relationship

ABSTRACT

Following a brief research review regarding the relationship between teacher behavior and student outcomes, a model is proposed for identifying those teaching behaviors that are significantly related to high-quality student performance. The model's stages include: (1) delineation of questions; (2) establishment of a framework; (3) selection of an empirical model; (4) selection of instrumentation; (5) development and validation of instruments; (6) organizational diagnosis; (7) teacher observation; (8) data collection, tabulation, and analysis; (9) interpretation of findings; (10) communication of results; (11) replication; and (12) refinement. An example is presented of the use of such a model to determine the effectiveness of secondary school mathematics teachers in a hypothetical school district. (CB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

9-10-87

ED287842

APPRAISING TEACHER PERFORMANCE:

A QUANTITATIVE APPROACH

BY

James G. Wingate, Ph.D.

Fred Bowers, Ph.D.

September 1987

SP029496

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

J. Wingate

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

INTRODUCTION

Discussions of teacher performance, student outcomes and a literate America are quite prevalent today. Moreover, guidelines of operation appear to have become more stringent and demands of accountability more forceful (Darling-Hammond, Wise and Pease, 1983). The Secretary of Education has delivered a number of national report cards on the status of American education and even though there is evidence of improvement, we still lag behind many of the less developed countries in terms of literacy.

Is there a single variable or a set of variables that appear to be the perpetrator of this less than excellent educational climate? While we can not target a single cause, we do know that a literate and informed populace, a trained labor force and citizens who demonstrate concern for their fellow humans have long been associated with effective teaching and learning.

A 1979 Gallop Poll revealed that the public overwhelmingly supported the improvement of the quality of teachers. Suggestions for improvement ranged from effecting changes in certification and tenure to the establishment of professional standards and practice boards.

The omnipresence of concerns for educational quality and reform gives rise to a number of questions. Two primary questions are as follows:

1. What constitutes a comprehensive and objective evaluation of teacher performance?

2. How may we relate teacher performance to student outcomes?

Before attempting to respond to these questions, one should cite those limitations and/or assumptions that are operable. The performance appraisal task is limited by the availability of models, the lack of quality instruments and the generalization of the results. Observer bias and insufficient sampling of performance also hamper the appraisal tasks. Further, supervisory ratings have generally been found to lack interrater reliability and validity (Medley, Peterson and Kauchak, 1982).

"Classroom observations are a common way of appraising teacher performance. They tend to reveal a view of the climate, rapport, interaction and functioning of the classroom available from no other source" (Evertson, and Holley, 1981). Darling-Hammond, Wise and Pease (1983) stated that to the extent that observations can capture what the teacher knows and does in interaction with a class of students, they can result in measures of performance.

There are a number of assumptions that appear to be consistent with the context and scope of teacher evaluation. Regardless of the validity or lack there of, these assumptions tend to enhance the tasks involved or exacerbate the problems associated with objective appraisal of teacher performance. Examples of these assumptions include but are not limited to the following:

1. Supervisors conduct an appropriate organizational diagnosis before embarking upon the performance appraisal of the teachers they supervise.

2. Periodically teachers are observed and evaluated.
3. The frequency with which supervisors observe teachers varies.
4. The observation/evaluation task is not void of certain biases and/or prejudices.
5. Supervisors are interested in assessing a number of common variables.
6. These variables may or may not be related to student achievement.

One can discern a number of nuances related to the appraisal of teacher performance. In the next section, teacher performance will be linked to student outcomes.

TEACHER PERFORMANCE AND STUDENT OUTCOMES

What do we really know about teacher behaviors and student outcomes? Rosenshine (1970) advanced the notion that teaching acts themselves may be unstable. Rosenshine and Furst (1971) and Heath and Nielsen (1974) concluded that most teaching acts are unrelated to student outcomes.

Dempsey-Atwood and Shavelson (1976) stated that teacher effects on pupil outcomes are unstable. Erlich and Shavelson (1978) agreed that most of the correlations between measures of teacher behavior and student outcomes have been low or nonreplicable. The accounts of instability among teacher behaviors and student outcomes have been countered by Stallings (1977) who stated that discrete sets of behaviors consistently lead to increased student performance. Darling-Hammond and Wise (1981) proposed a link between teacher performance and student achievement. Support for this proposal was found in 1983, when Darling-Hammond, Wise and Pease concluded that there is little evidence that single teaching variables can be identified as essential for effective teaching. Further they stated that differences in patterns of teaching performances contribute to learning and that these patterns are not uniformly applicable to all grade levels, subject areas and teaching.

This line of reasoning is consistent with Land & Smith's (1979) differentiation between high-inference teacher clarity variables (open to subjectivity) and low-inference teacher clarity variables (objectively quantifiable). Examples of high-inference

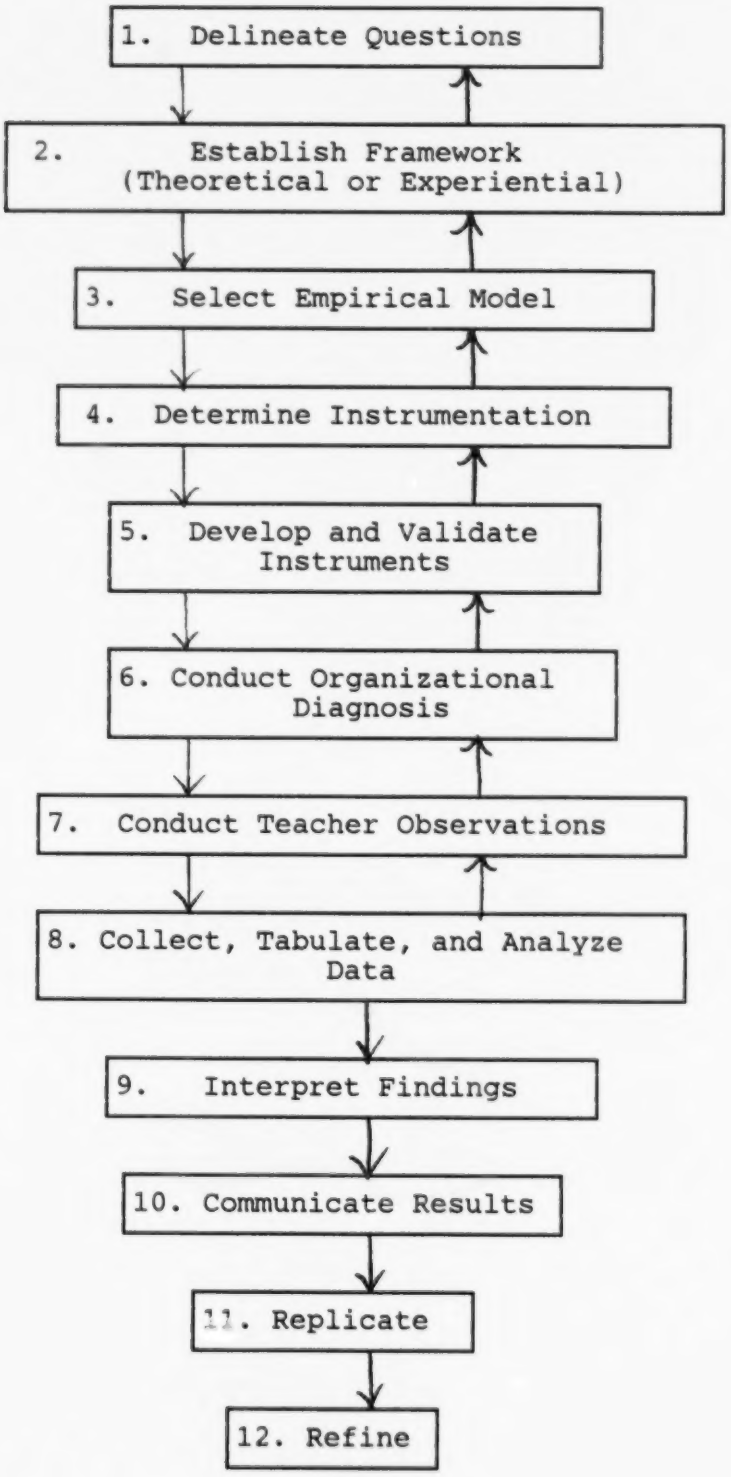
variables include clarity, variability, enthusiasm, task oriented or business-like behaviors, student opportunity to learn criterion material, use of student ideas, general indirectness, use of structuring comments, criticism, types of questions and probing (Centra and Potter, 1980); while bluffing and lesson discontinuity are common low inference variables. Smith (1985) found that the presence of these variables on the part of high school social studies teachers affected student achievement in a negative fashion.

Based on the aforementioned accounts, one can conclude that measurable teaching behaviors are generally unstable. This instability among variables limits the reliability with which one can predict student outcomes. It is imperative then that one should operationally define the variables under study and ensure that they are relevant to the specific investigation (correlation or other statistical treatments). A quantitative approach that addresses the following questions would be appropriate:

1. What is the state of the research as it relates to teacher evaluation?
2. How do we define teacher effectiveness?
3. Which variables should we investigate?
4. Which variables explain the variance between those instructors who are considered more effective than others?

The following model lends itself to addressing the questions raised above:

A MODEL FOR
PERFORMANCE APPRAISAL OF TEACHERS



Before proceeding, one should conduct a review of the current literature on teacher evaluation particularly as it relates to the questions that have been delineated. From this review, one should be in a position to suggest a framework or a theoretical position with a built-in rationale (Step 2).

Selection of an empirical model for one's investigation (Step 3) should be consonant with the selection and validation of the instruments to be utilized in data collection (Steps 4 & 5). An instrument for diagnosing the climate and well-being of ones organization should be included (Step 6). Assessing the validity and reliability of the selected instruments is a critical task that should not be overlooked.

Teacher observations should be conducted; data should be collected, tabulated and analyzed, and the findings interpreted. Painstakingly efforts should be taken to communicate the results to all parties concerned in a diplomatic and objective manner. Conclusions and inferences should be cautiously drawn and replication should be made for the purpose of assessing consistency and reliability. The model should be refined in accordance with any new findings.

In the following section, a scenario is developed for the purpose of demonstrating how one might begin to address the question of effectiveness among mathematics instructors. Even though the example is one demonstrative of a high school situation, it could be applicable to other secondary and post-secondary situations.

THE CASE OF THE "ANYWHERE" SCHOOL DISTRICT

Through a series of interviews with the Superintendent, mathematics instructors, board members and members of the general public, the mathematics supervisor concluded the following:

1. The level of satisfaction with the quality of mathematics instruction was low. In general, the effectiveness of mathematics instructors was questionable.
2. Teachers in general were thought to be less than effective.
3. The frequency and quality of teacher observations should be addressed.
4. Evidence of improvement of quality should be shown annually in the form of a larger percentage of students scoring above the mean ($\bar{x}$) that is set by the district.
5. Evidence of increased effectiveness on the part of instructors should be shown in terms of the manner in which they plan, organize, and implement the curriculum.
6. The required paper work infringed greatly on teachers' preparation time.

The supervisor proceeded to invoke her best thinking along with creativity and ingenuity. This resulted in the following problem statement and related questions:

Problem Statement

What relationship(s) exist between teacher behaviors and student outcomes in the "Anywhere" School District?

Related Questions

1. What are the attributes of an effective teacher in the district?
2. Which variables explain the variance between those instructors who are rated effective and those who are rated less effective.
3. Are planning skills on the part of teachers significantly different between effective and less effective instructors?

A performance appraisal instrument was developed, validated and then used to assess the quality of teaching. The variables that were studied were categorized as follows: planning, organizing, implementing, managing, monitoring and supervision (See Appendix A). Teachers were then asked to provide a distribution of both raw and standard scores for their classes. The two sets of data were correlated and the results are shown in Appendix B.

Note that student scores were transformed into standard scores (t distribution with a mean of 50 and a standard deviation of 10). Further a $\underline{t}$ score $\geq$ 55 was selected as the criterion of acceptability. The percentage of criterion scores for a given class was correlated with the observation score of the teacher of the class.

The overall results was a Pearson r (correlation coefficient) of .53. In general, a correlation of .53 is considered to be moderate. In this particular case with 10 samples and 8 degrees of freedom, we can be between 80% and 90% confident that the correlation was not due to chance alone (see Appendix C).

The mathematics supervisor should interpret the results with extreme caution. The interpretation should be as follows:

> There appears to be a significant relationship between certain observed teacher behaviors and student outcomes. In order to assess the stability of this finding, it is suggested that the study be replicated and the results compared. If the findings are comparable, the possibility of moving to a more comprehensive statistical design should be investigated (analysis of variance, factor analysis, or multiple regression).

Summary and Implications

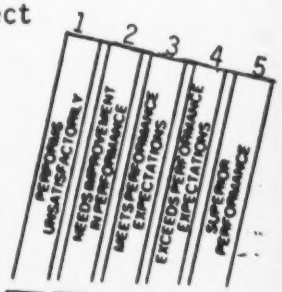
At the very least, one can conclude that there is disagreement among researchers with regards to the relationships among teacher behavior and student outcomes. To the extent that we can identify those teaching behaviors that are significantly related to quality student performance and quantify these accordingly, we can begin to distinguish between style and substance. What may appear to be good educational practices on the surface may in effect produce little in the way of significant changes in student performance. This suggests that efforts to relate teacher behaviors and student outcomes should continue. Further, a knowledge of strategies with impact can become invaluable in the task of allocating limited resources.

ANYWHERE SCHOOL DISTRICT

TEACHER PERFORMANCE APPRAISAL INSTRUMENT

Instructions: Based on evidence that has been derived from observation and discussion, rate the teachers performance with respect to the attributes listed below.

Please check one box per attribute.



A. Planning

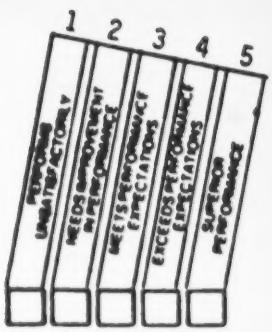
1. Contributes to the development of annual objectives for the school. ☐☐☐☐☐
2. Develops an annual instructional plan that is consistent with the annual objectives for the school. ☐☐☐☐☐
3. Anticipates outcomes and develops contingency plans accordingly. ☐☐☐☐☐
4. Presents information in a timely and understandable manner. ☐☐☐☐☐

B. Organizing

5. Utilizes students as a resource in instructing, developing materials and operating equipment. ☐☐☐☐☐
6. Utilizes volunteers, professionals, and other community resources for the purpose of extending classroom learning. ☐☐☐☐☐
7. Enhances learning activities by incorporating field trips. ☐☐☐☐☐

C. Implementing

8. Utilizes a variety of instructional methods. ☐☐☐☐☐
9. Utilizes plans and activities that allow for individual differences among students. ☐☐☐☐☐
10. Provides timely, objective, and accurate feedback to students and parents. ☐☐☐☐☐
11. Fosters and maintains control of the classroom environment. ☐☐☐☐☐



C. Implementing (Continued)

12. Encourages participation on the part of students. ☐☐☐☐☐

13. Exhibits patience and understanding with students. ☐☐☐☐☐

D. Managing and Monitoring

14. Monitors pupils' progress daily. ☐☐☐☐☐

15. Ensures proper balance of activities, lecture and hands-on experiences. ☐☐☐☐☐

16. Develops and utilizes a variety of assessment and evaluation instruments. ☐☐☐☐☐

17. Maintains appropriate records and documentation. ☐☐☐☐☐

18. Utilizes formative and summative techniques for evaluative purposes. ☐☐☐☐☐

19. Holds students accountable for their actions. ☐☐☐☐☐

20. Maintains consistency and objectivity in grading practices. ☐☐☐☐☐

E. Supervision

21. Utilizes time and resources effectively. ☐☐☐☐☐

22. Deals directly with individual conflict and disruptive behavior. ☐☐☐☐☐

23. Maximizes positive reinforcement and maintains a climate that maximizes learning. ☐☐☐☐☐

24. Participates in staff development activities. ☐☐☐☐☐

25. Carries out other assigned duties in a positive fashion. ☐☐☐☐☐

Score Distribution: Math Classes

SMITH ALGEBRA I SCORES	T SCORE	JOHNSON ALGEBRA I SCORES	T SCORE	WHITE ALGEBRA I SCORES	T SCORE
94	65	95	67	95	66
93	64	94	66	95	66
93	64	90	63	95	66
90	61	90	63	90	62
90	61	87	60	90	62
89	60	87	60	89	61
87	58	85	58	87	59
87	58	85	58	87	59
85	56	85	58	87	59
85	56	85	58	85	57
85	56	84	57	85	57
84	55	84	57	84	56
84	55	82	55	82	54
84	55	80	53	82	54
82	53	80	53	82	54
82	53	80	53	77	50
80	51	77	50	77	50
77	49	75	48	77	50
77	49	75	48	77	50
75	47	75	48	75	48
75	47	71	45	71	44
68	40	71	45	71	44
66	38	71	45	68	42
66	38	71	45	68	42
65	37	66	40	68	42
65	37	65	39	66	40
65	37	65	39	65	39
64	36	64	38	65	39
60	33	60	34	65	39
58	31	60	34	64	38
		58	32	60	34
		58	32	60	34
				58	33

MEAN	78.5	76.7	77.2
STD DEV	10.63876	10.58960	10.98884

CRITERION IS 55	# TO MEET OR EXCEED	# TO MEET OR EXCEED	# TO MEET OR EXCEED
	14	13	12
	% TO MEET OR EXCEED	% TO MEET OR EXCEED	% TO MEET OR EXCEED
	46.7%	40.6%	36.4%

TALARICO ALGEBRA I SCORES	T SCORE
98	65
96	63
96	63
95	63
95	63
95	63
93	61
93	61
93	61
90	58
87	56
86	55
86	55
85	54
85	54
85	54
83	52
82	52
80	50
80	50
77	47
75	46
75	46
73	44
73	44
70	41
68	40
68	40
67	39
66	38
66	38
64	36
64	36
60	33
59	32

ASNER ALGEBRA I SCORES	T SCORE
100	70
96	67
94	65
94	65
90	62
90	62
90	62
87	59
85	57
85	57
84	56
84	56
84	56
83	56
79	52
79	52
79	52
77	50
77	50
75	49
75	49
75	49
73	47
69	43
69	43
69	43
68	43
67	42
66	41
66	41
65	40
64	39
64	39
60	36
59	35
57	33
57	33

SODINSKI ALGEBRA I SCORES	T SCORE
98	67
98	67
96	65
95	64
94	63
94	63
94	63
90	60
90	60
89	59
86	57
85	56
85	56
84	55
84	55
84	55
84	55
83	55
80	52
79	51
79	51
75	48
73	46
73	46
70	44
70	44
69	43
68	42
68	42
68	42
67	41
67	41
66	41
65	40
65	40
64	39
60	36
60	36
60	36
59	35
57	33

MEAN	STD DEV
80.2	11.69147

MEAN	STD DEV
76.6	11.54392

MEAN	STD DEV
77.4	12.27340

	# TO MEET OR EXCEED	% TO MEET OR EXCEED
CRITERION IS 55	13	37.1%
	14	37.8%
	18	43.9%

BROWN ALGEBRA I SCORES	T SCORE	MOSS ALGEBRA I SCORES	T SCORE	STEINBERG ALGEBRA I SCORES	T SCORE
95	72	98	70	97	65
90	67	95	68	97	65
90	67	95	68	94	63
88	64	95	68	94	63
85	61	90	63	94	63
85	61	90	63	90	60
85	61	86	59	90	60
82	58	86	59	87	57
82	58	82	55	87	57
80	56	80	53	87	57
78	54	80	53	85	56
78	54	80	53	85	56
75	51	78	51	85	56
75	51	78	51	83	54
75	51	78	51	83	54
72	47	75	48	79	51
72	47	75	48	79	51
70	45	75	48	79	51
70	45	75	48	72	45
70	45	73	47	72	45
70	45	73	47	72	45
70	45	70	44	72	45
68	43	70	44	69	43
68	43	70	44	65	40
68	43	68	42	62	37
65	40	68	42	62	37
65	40	65	39	60	36
65	40	65	39	60	36
65	40	65	39	60	36
60	35	65	39	59	35
60	35	65	39	59	35
60	35	60	34		
		60	34		

MEAN	74.4	MEAN 76.6	MEAN 78.0
STD DEV	9.377031	STD DEV 10.47056	STD DEV 12.47056

	# TO MEET OR EXCEED	# TO MEET OR EXCEED	# TO MEET OR EXCEED
CRITERION IS 55	10	9	13
	% TO MEET OR EXCEED 31.3%	% TO MEET OR EXCEED 27.3%	% TO MEET OR EXCEED 41.9%

SHEA ALGEBRA I SCORES	T SCORE
97	66
95	64
95	64
94	64
94	64
93	63
90	60
89	59
89	59
87	57
87	57
87	57
85	56
85	56
84	55
82	53
82	53
82	53
80	51
77	49
77	49
77	49
75	47
75	47
71	44
68	41
68	41
68	41
68	41
66	39
66	39
65	38
65	38
64	37
60	34
58	32
58	32

PERFORMANCE CORRELATION - PEARSON r

X	Y	X^2	Y^2	XY
122	47	14884	2178	5693
108	41	11664	1650	4388
120	36	14400	1322	4364
95	37	9025	1380	3529
115	38	13225	1432	4351
120	44	14400	1927	5268
104	31	10816	977	3250
110	27	12100	744	3000
120	42	14400	1759	5032
120	41	14400	1644	4865
1134	384	129314	15012	43740

n = 10

PEARSON r = 0.53

MEAN
78.5

STD DEV
11.50697

# TO MEET OR EXCEED	% TO MEET OR EXCEED
CRITERION IS 55	15	40.5%

Table Of "r_{test}" Values -- Test Values for Correlation Coefficient

# of Samples	(df) degrees of Freedom	80%	90%	95%	99%	99.9%
3	1	.951	.988	.997	1.000	1.000
4	2	.800	.900	.950	.990	.999
5	3	.687	.805	.878	.959	.991
6	4	.608	.729	.811	.917	.974
7	5	.551	.669	.755	.875	.951
8	6	.507	.621	.707	.834	.925
9	7	.472	.582	.666	.798	.898
10	8	.443	.549	.632	.765	.872
11	9	.419	.521	.602	.735	.847
12	10	.398	.497	.576	.708	.823
13	11	.380	.476	.553	.684	.801
14	12	.365	.457	.532	.661	.780
15	13	.351	.441	.514	.641	.760
16	14	.338	.426	.497	.623	.742
17	15	.327	.412	.482	.606	.725
18	16	.317	.400	.468	.590	.708
19	17	.308	.389	.456	.575	.693
20	18	.299	.378	.444	.561	.679
21	19	.291	.369	.433	.549	.665
22	20	.284	.360	.423	.537	.652
23	21	.277	.352	.413	.526	.640
24	22	.271	.344	.404	.515	.629
25	23	.265	.337	.396	.505	.618
26	24	.260	.330	.388	.496	.607
27	25	.255	.323	.381	.487	.597
28	26	.250	.317	.374	.479	.588
29	27	.245	.311	.367	.471	.579
30	28	.241	.306	.361	.463	.570
31	29	.237	.301	.355	.456	.562
32	30	.233	.296	.349	.449	.554
42	40	.202	.257	.304	.393	.490
62	60	.165	.211	.250	.325	.408
122	120	.117	.150	.178	.232	.294

The values in this table are from the formula:

$$r_{test} = \frac{(\quad t^2 \quad)\ 1/2}{(t^2 + \ df)}$$

(Texas Instruments, TI-55 III Guidebook)

REFERENCES

COOLEY, W.W. and LOHNES, P.R. Evaluation Research in Education, New York: Irvington Pub., Inc., 1976.

CRONBACK, L. J., GLESER, G. C., NANDA, H. and RAJARATNAM, N. "The Dependability of Behavioral Measurements: The Theory of Generalizability For Scores and Profiles." New York: John Wiley and Sons, 1972.

DARLING-HAMMOND, L., WISE, A.E., & PEASE, S.R. "Teacher Evaluation in the Organizational Context: A Review of the Literature," Review of Educational Research, Volume 53, Number 3, 285-328.

DOWNIE, N.M. & HEATH, R.W. Basic Statistical Methods, New York Harper & Row Pub., 1974.

ERLICK, O., & SHAVELSON, R. Generalizability of Measures: A Computer Program For Two- and Three-facet Designs. "Behavior Research Methods and Instrumentation." 1976, 8, 407-409.

GRONLUND, N.E. Measurement and Evaluation in Teaching, New York: The Macmillan Co, 1971.

KERLINGER, F. N. & PEDHAZUR, E.J. Mulitple Regression in Behavioral Research, New York: Holt, Rhinehart and Winston, Inc. 1973.

ROSENSHINE, B. The Stability Of Teacher Effects Upon Student Achievement. "Review of Educational Research." 1970, 40, 647-662.

ROSENSHINE, B. & FURST, N. Research on teacher performance criteria. In B. O. Smith (ed.), "Research In Teacher Education." Englewood Cliffs, New Jersey: Prentice-Hall, 1971.

ROSENSHINE, B. & FURST, N. The Use Of Direct Observation To Study Teaching. In. R. Travers (Ed.), "Second Handbook Of Research On Teaching." Chicago: Rand McNally, 1973.

SANDOVAL, J. Beginning Teacher Evaluation Study Completion Report, Task K: Sub-study Of Consistency Of Teacher Behavior. Princeton, N.J.: Educational Testing Service, 1974.

SHAVELSON, R.J., & DEMPSEY-ATWOOD, N. Generalizatility of measures of teaching behavior. "Review of Educational Research". 1976, 46, 553-611.

SMITH, L.R. "Presentational Behaviors and Student Achievement in Mathematics," Journal of Educational Research, Volume 78, Number 5, (1985): 292-297.

SMITH, L.R. "Teacher Clarifying Behaviors: Effects on Student Achievement and Perceptions," Journal of Experimental Education, Volume 3 (1985): 162-169.

WALBERG, H.J. "Models for Optimizing and Individualizing School Learning," Interchange, Volume 2, Number 3 (1971): 15-27.

WALBERG, H.J. Evaluating Educational Performance, California: McCutchan Publishing Corporation, 1974.